Classification of EEG Signals Based on Image Representation of Statistical Features

Chapter · January 2020

DOI: 10.1007/978-3-030-29933-0_37

4 authors:

Jodie Ashford
Aston University

Jordan J. Bird
Aston University

Felipe Campelo
Aston University

Diego R. Faria
Aston University

Classification of EEG Signals Based on Image Representation of Statistical Features

Jodie Ashford*, Jordan J. Bird*, Felipe Campelo, and Diego R. Faria

School of Engineering and Applied Science
Aston University, Birmingham, UK
{ashfojsm, birdj1, f.campelo, d.faria}@aston.ac.uk

Abstract. This work presents an image classification approach to EEG brainwave classification. The proposed method is based on the representation of temporal and statistical features as a 2D image, which is then classified using a deep Convolutional Neural Network. A three-class mental state problem is investigated, in which subjects experience either relaxation, concentration, or neutral states. Using publicly available EEG data from a Muse Electroencephalography headband, a large number of features describing the wave are extracted, and subsequently reduced to 256 based on the Information Gain measure. These 256 features are then normalised and reshaped into a 16×16 grid, which can be expressed as a grayscale image. A deep Convolutional Neural Network is then trained on this data in order to classify the mental state of subjects. The proposed method obtained an out-of-sample classification accuracy of 89.38%, which is competitive with the 87.16% of the current best method from a previous work.

Keywords: Machine Learning, Convolutional Neural Networks, Image Recognition, Mental State Classification, Electroencephalography

1 Introduction

Human-machine interaction is often considered a mirror of the human experience; sound and visuals constitute voice recognition, human activity classification, facial recognition, sentiment analysis and so on. Though, with the availability of sensors to gather data that the human body cannot, interaction with machines can often exceed the abilities of the natural human experience. An example of this is the consideration of electroencephalographic brainwaves. The brain, based on what a person is thinking, feeling, or doing, has a unique pattern of electrical activity that emerges as a consequence of the aggregate firing patterns of billions of individual neurones [1, 2]. These electrical signals can, in principle, be detected and processed to infer the state of the brain and, by extension, the mental state of a given subject. Besides clinical applications, this possibility is also useful, e.g., for brain-machine interfacing.

* JA and JJB are co-first authors.

More effective methods of feature extraction and classification are of utmost importance in brain-machine interaction, since better performing models can interpret human brain activity with higher accuracy. Previous works [3, 4] suggest that static statistical descriptions of brainwaves present the information in the signals in a more machine learning-friendly shape than the raw waves themselves, even when temporally-aware machine learning methods are employed.

This work focuses on the process of feature extraction, selection and formatting in order to achieve improved classification accuracy of EEG signals. More specifically, the main contribution is a framework to perform classification of these signals, based on (i) the extraction of a large number of static statistical features of the data, followed by (ii) automated feature selection and (iii) representation of the selected attributes as a 2D matrix. The resulting matrices are (iv) interpreted as grayscale images, which allows the leveraging of the state-of-the-art performance of convolutional neural networks [5, 6] as image classifiers.

The remainder of this paper is organised as follows. A brief presentation of the background concepts related to the present work is provided in Section 2, followed by the description of the proposed approach in Section 3. The results obtained by the proposed method are discussed in Section 4. Finally, conclusions and suggestions of further investigations are provided in Section 5.

2 Background

Electroencephalography (EEG) is a technique used to measure the electrical activity of a brain. The human brain contains billions of neurones, which each exhibit electrical activity in the form of nervous impulses [7, p. 31]. The electrical signal produced by a single neurone is difficult to detect, but the combined signal from the action of many neurones together can be measured using EEG [8, p. 4].

Typically, EEG involves placing electrodes onto the scalp of the subject. These electrodes measure the voltage fluctuations generated by thousands of active neurones in the brain. These signals are then digitised and amplified [9, 10]. Possibly the main advantages of this method of measuring brain activity are that it is a non-invasive and inexpensive technique. Even less invasive techniques, such as the Muse headband, have extended the utility of EEG beyond medical examination alone, at the expense of sensitivity. Unlike imaging techniques such as MRI, EEG can measure fluctuations in electrical activity on the scale of milliseconds, which makes it an incredibly powerful tool for measuring real-time brain activity in response to stimulus [10].

The ability to infer human mental states is as important in human-machine interaction as a form of Affective Computing[11] as it is in natural human interaction. In the past, such techniques have used attributes available to humans: speech, gestures, facial expressions, etc. [12, 13]. With the increasing development of non-invasive EEG technology, researchers can take advantage of sensors available only to machines to attempt to classify human emotions directly from the brain. Such analysis is less dependent on environmental factors or differences in somatic expression between individuals. It also offers a more seamless avenue of human-machine communication.

2.1 Related Work

Non-invasive EEG headsets have been used in previous works to analyse mental states. In a related example, data from the Muse headset was was found to be useful in the evaluation of the enjoyment levels of subjects playing two different video games [14]. The findings aligned with the current understanding of how waves detected by EEG (in this case, frontal theta frequencies) map to enjoyment. This is an example of how non-invasive techniques can provide uses of EEG outside of the medical setting, and provide data for emotional classification.

Previous works have also shown the excellent performance of convolutional neural networks (CNN) in EEG-based mental state classification. In 2017, using the DEAP dataset [5], EEG signals were classified using both deep (DNN) and convolutional (CNN) neural networks [6]. Two different classifications were performed: one for valence and one for arousal, classifying each as either high or low. The DNN achieved 75.78% accuracy for valence and 73.28% for arousal, while the CNN achieved 81.41% and 73.35% respectively [6].

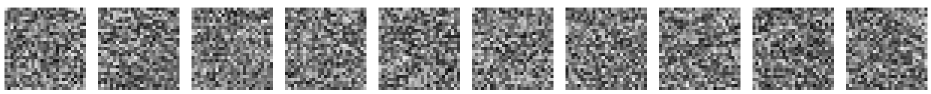
Projection of EEG data onto a "visual" space is a fairly recent approach, with relatively little work as of yet performed into its exploration. Most of the relevant literature in this area [15, 16] relates to mapping the signal readings of electrodes to a spatial representation of the brain itself, interpolating intermediary points based on values from the nearest electrodes. Alternatively, some limited but successful work has explored the CNN classification of visual spectrograms produced by the signals[17]. Spectrograms produced by Limited Field Potentials have also found varying levels of success in classifying biological signals from rat brains[18, 19]. In these works, a limited set of five features were extracted and machine learning approaches (decision trees, discriminant analysis, support vector machines, and nearest neighbour classifiers) were used to recognise patterns, producing results with accuracies ranging from 95.8% to 98.8%. These solutions, though effective, rarely consider statistical processing of the waves as a way to extract relevant data from the complex waveforms generated by EEG. Visual pixel-wise approaches and subsequent CNN applications have been successfully implemented in other biological domains such as image segmentation of electron microscopy images [20], with promising results for a variety of applications.

The solution suggested in this study, on the other hand, is based on extracting statistical features from EEG signal waves and maps them onto static 2D matrices, which are then represented as images and used for the classification of mental states using a convolutional neural network. This proposed methodology is detailed in the following section.

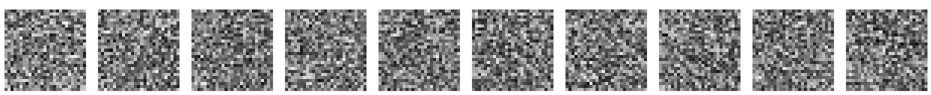
3 Proposed Approach

Firstly, an available training set of EEG signals is preprocessed. The data is assumed to contain the time series related to one or more electrodes, within a given experimental time frame, labelled in terms of three distinct mental states (relaxed, concentrating, and neutral) that the subjects were keeping during data collection [4]. From these signals a number of statistical features are extracted [3, 4], resulting in a high dimensional attribute space - in the case of this work, 1274 features are generated for each time window, as detailed in Section 3.1. To focus only on the most relevant ones for the classification process, feature selection is applied to the resulting features. Here, the $16^2 = 256$ most descriptive ones, based on the estimated information gain [21], are selected.

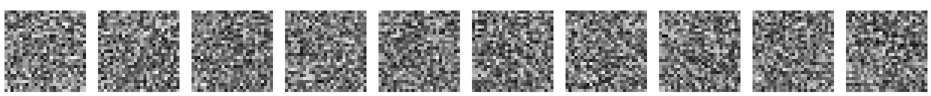
Finally, the selected features are converted into a 16×16 grid of numerical values normalised to the $[0, 1]$ range, which can be represented as a grayscale image. Figure 1 shows a number of samples of relaxed, neutral and concentrating brainwave data, using this particular image representation.



(a) Ten samples of relaxed brainwave data represented as 16×16-pixel images



(b) Ten samples of concentrating brainwave data represented as 16×16-pixel images



(c) Ten samples of neutral brainwave data represented as 16×16-pixel images

Fig. 1: Examples of image representations for each of the three mental states considered in this work.

The resulting set of images is then used to train a convolutional neural network (CNN) [22] as a classifier of the three mental states investigated in this particular work. The details of the CNN are provided in Section 3.2.

3.1 Feature Extraction

Due to the temporal, auto-correlated nature of the EEG waves, single-point features cannot generally provide enough information for good rules to be generated by machine learning models. In this work we follow the approach of extracting statistical features based on sliding time windows [3, 4]. More specifically, the EEG signal is divided into a sequence of windows of length one second, with consecutive windows overlapping by 0.5 seconds, e.g., $[(0s - 1s), [0.5s - 1.5s), [1s - 2s), \ldots])$.

Assume that each 1-second time window contains a sequence $\mathbf{x} = [x_1, \ldots, x_N]$ composed of N samples. Also let $\mathbf{x}_{h1}$ and $\mathbf{x}_{h2}$ denote the first and second halves of the window, and $\mathbf{x}_{q1}, \mathbf{x}_{q2}, \mathbf{x}_{q3}, \mathbf{x}_{q4}$ denote the four quarter-windows obtained by dividing the window into four (roughly) equal-sized parts, each composed of approximately $N/4$ samples.[1]

In this work the following statistical features were generated for each time window:

- Considering the full time window:
  - The sample mean and sample standard deviation of each signal (8 features).
  - The sample skewness and sample kurtosis of each signal [23] (8 features).
  - The maximum and minimum value of each signal (8 features).
  - The sample variances of each signal, plus the sample covariances of all signal pairs [24] (10 features).
  - The eigenvalues of the covariance matrix [25] (4 features).
  - The upper triangular elements of the matrix logarithm of the covariance matrix [26]. (10 features)
  - The magnitude of the frequency components of each signal, obtained using a Fast Fourier Transform (FFT) [27] (300 features).
  - The frequency values of the ten most energetic components of the FFT, for each signal (40 features).
- Considering the two half-windows:
  - The change in the sample means and in the sample standard deviations between the first and second half-windows, for all signals (8 features).
  - The change in the maximum and minimum values between the first and second half-windows, for all signals (8 features).
- Considering the quarter-windows:
  - The sample mean of each each quarter-window, plus all paired differences of sample means between the quarter-windows, for all signals (56 features).
  - The maximum (minimum) values of each quarter-window, plus all paired differences of maximum (minimum) values between the quarter-windows, for all signals (112 features).

Regarding the representation of the signals in the frequency domain using FFT [27], two specific aspects were taken into account: first, the DC-component of the signals was filtered out prior to the application of the FFT, so the zero-frequency component was always set as zero. This was done to prevent the offset to completely dominate the power spectrum, even though it carries no relevant information for the classification task. The second aspect is that frequencies in the range of (50 ± 1) Hz were also filtered out, to remove any contamination from the AC electrical distribution frequency, which could also skew the power spectrum of our signals.

[1] In this work we standardised the number of samples within each window to $N = 150$. This means that quarter-windows have either $n = 37$ or $n = 38$ observations.

Each window receives as features the vector of quantities computed above for both itself and the window that immediately precedes it (*1-lag window*). Features from the 1-lag window that were clearly redundant due to the half-window overlaps were removed prior to the composition of the feature vector, namely the sample means, maximum and minimum values of $\mathbf{x}_{q3}$ and $\mathbf{x}_{q4}$, as well as their respective differences. In the end a total of 989 features were generated for each time window (except the first, which was only used as the 1-lag for the second one).

After the statistical features were extracted the resulting dataset was composed of 2479 data objects, each represented by its corresponding 989 feature values plus a single class label. Feature selection was then performed based on the Information Gain of each feature, and the total number of features was reduced to 256 (plus class label). Due to privacy considerations the raw EEG data cannot be released, but the processed dataset is publicly available at https://www.kaggle.com/birdy654/eeg-mental-state-v2 as a UTF-8 encoded CSV with approximately 6MB.

3.2 Convolutional Neural Network

Convolutional neural networks (CNN) [28] are a specialised kind of neural network for processing data that has a known grid-like topology, which makes them particularly suitable for dealing with data represented as time series or images [22]. The main distinguishing feature of these networks is the use of a convolution [29] instead of simple matrix multiplication in at least one of their layers [22]. Convolutional neural networks are generally very effective at image classification tasks [30–32], which motivates their use here. For more details on these networks, please refer to Ian Goodfellow *et al.*'s book on the subject [22].

In this particular work we have opted for using the CNN implementation available in the Keras Deep Learning Python library [33]. The network was trained on an Nvidia GTX1060 (1280 CUDA Cores, 6GB 8Gbps GDDR5 VRAM). The topology and hyperparameters of the convolutional neural network were defined based on preliminary, trial-and-error experimentation. Table 1 shows the resulting model for the classification of brainwave images.

Other design choices that were arbitrarily set in this experiment are the use of the ADAM optimiser [34] to train the network; and the use of a batch size of 100, trained for 400 epochs, with the loss calculated via categorical cross entropy at a 70/30 validation split:

$$CE = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}), \tag{1}$$

where M is the number of class labels (in this case, 3), y is a binary indication of a correct prediction (1 or 0), and p is the predicted probability of observation o of class c. The entropy of each class within the testing split is calculated and added for a final, overall result. In this case, this is the entropy of the three classes of mental state - relaxed, neutral, and concentrating.

Table 1: Network topology and parameters used. Please refer to the Keras documentation [33] for specific definitions.

Layer	**Output**	**Params**
Conv2d (ReLu)	(0, 14, 14, 32)	320
Conv2d (ReLu)	(0, 12, 12, 64)	18496
Max Pooling	(0, 6, 6, 64)	0
Dropout (0.25)	(0, 6, 6, 64)	0
Flatten	(0, 2304)	0
Dense (ReLu)	(0, 512)	1180160
Dropout (0.5)	(0, 512)	0
Dense (Softmax)	(0, 3)	1539

4 Results

In this section the results for the experiments are presented. The experiments were performed three times, the difference between the three runs being random seeds set at the start of the experiment. The overall final score always resulted within 400 epochs. Accuracy and loss per-epoch are illustrated for the first run.

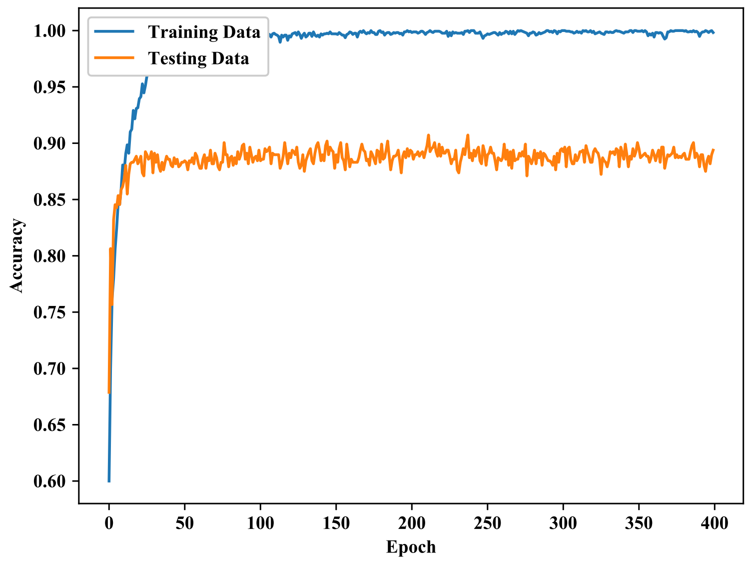
4.1 Results obtained

Figure 2 illustrates the accuracy and loss of the network, for both training and testing data from the validation split. The overall out-of-sample accuracy of the CNN in classifying the dataset was 89.38% (665/744 correct classifications, $CI_{0.95} = [86.94, 91.50]\,\%$). As can be observed, the accuracy curve saturates after about 50 epochs, after which the loss starts increasing. This suggests that computational resources are essentially wasted after this point, and more parsimonious training can be employed in the future.

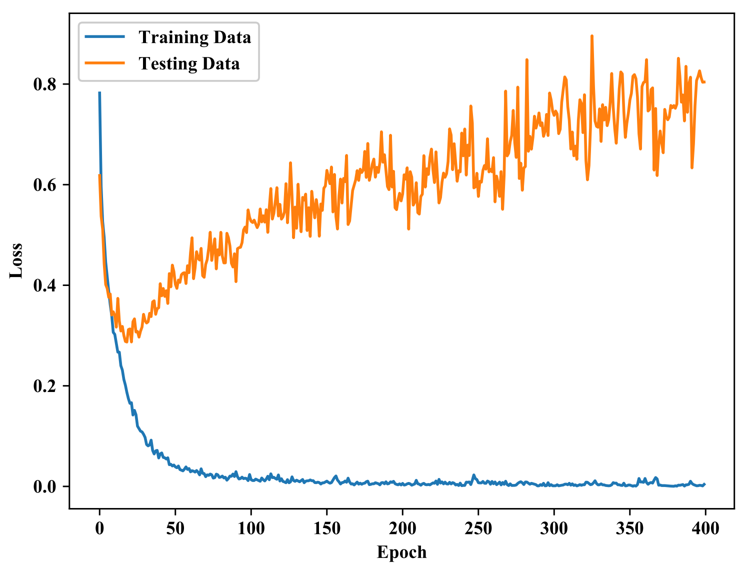
Table 2: Comparison with Related Studies using the Same Dataset as this Experiment. Column *Accuracy* also provides 95% confidence intervals for the accuracy.

Study	**Method**	**Validation**	**Focus**	**Accuracy**
This study	Inf. Gain Selection, CNN	70/30 Split	Accuracy	89.38% $[86.94, 91.50]$
[3]	OneR Selection, Random Forest	10-fold	Accuracy	87.2% $[85.7, 88.6]$
[35]	Evol. Selection, DEvoMLP	5-fold	Accuracy, Resource Usage	79.8% $[78.1, 81.5]$

Table 2 contrasts the results obtained in this paper with previous works dealing with the same mental state dataset. It is worth mentioning that only



(a)



(b)

Fig. 2: (a) Accuracy and (b) Loss of the CNN for Training and Testing Data across 400 epochs.

one of the compared experiments had the single goal of maximising accuracy, while the other was also focused on minimising computational effort. Another noteworthy point is that both previous works used cross-validation instead of a split set in order to estimate accuracy. With these factors in mind, the approach used in the present work has provided results that seem to be very competitive, with a point estimate of the accuracy that is approximately 2.18% greater (in absolute terms) than the one reported in the 2018 study. This difference is not, however, statistically significant at the 95% confidence level ($p = 0.129$ using the chi-squared test for equality of two binomial proportions [24]). Despite not clearly outperforming the current state of the art, this result suggests that the proposed approach of coupling an image-based representation of the data with CNN-based classification may represent an effective new strategy for performing EEG classification, with potential extensions to classification in the context of general time series data.

5 Future Work and Conclusions

In this work, a new approach for classification of EEG signals has been presented, based on the sequential application of statistical feature extraction and selection, normalisation and subsequent projection of the selected features as small images, and classification based on a convolutional neural network. The results obtained for this method have been shown to be very competitive with the best known results to date for the available dataset.

Possibly the most clear limitation of the present work is related to the question of generality. Since a single dataset is used, it remains to be seen how well the proposed methodology generalises not only to larger, possibly richer EEG data, but also - and more interestingly - to other similar time series. In this regard, further testing and statistical assessment of the proposed methodology are fundamental next steps as this line of research progresses.

Due to the limited available resources, the experiment reported in this work used a simple 70/30 data split instead of the more usual (but more computationally demanding) cross-validation, which should be used in future experiments whenever possible so as to obtain better estimates of out-of-sample accuracy[36]. Two other aspects related to the issue of limited resources were present. The first was the lack of a principled parameter tuning approach for both the structure and other parameters of the network, which can be optimised using, e.g., iterated racing [37], hyperheuristics [38], or topology-specific tuning methods [39–41]. Even under more constrained computational budgets, traditional design and analysis of experiments approaches [42] can be useful in defining the best network for this particular problem. The second issue is related with the selection of only 256 features to compose the image to be used in the training of the CNN. Future work in this direction should concern the testing of varying image sizes in order to better fine-tune the attribute selection process. In addition, further methods of feature extraction should be investigated and compared, rather than focusing solely on Information Gain as this study has done. The investigation of other

CNN architectures, which have shown much promise in other contexts [43], is also an interesting point for further development.

Regardless of the possible improvements discussed above, we argue that the proposed framework of projecting selected features onto a 2D matrix and subsequent image recognition through a Convolutional Neural Network already constitutes a competitive approach for brainwave data classification. The results obtained are promising, as compared to current scientific standards, and further exploration is strongly suggested to advance the results beyond the preliminary outcome presented in this paper.

References

1. R. Caton, "The electric currents of the brain," *American Journal of EEG Technology*, vol. 10, no. 1, pp. 12–14, 1970.
2. R. R. Llinás, "Intrinsic electrical properties of mammalian neurons and cns function: a historical perspective," *Frontiers in cellular neuroscience*, vol. 8, p. 320, 2014.
3. J. J. Bird, L. J. Manso, E. P. Ribiero, A. Ekart, and D. R. Faria, "A study on mental state classification using eeg-based brain-machine interface," in *9th International Conference on Intelligent Systems*, IEEE, 2018.
4. J. J. Bird, A. Ekart, C. D. Buckingham, and D. R. Faria, "Mental emotional sentiment classification with an eeg-based brain-machine interface," in *The International Conference on Digital Image and Signal Processing (DISP'19)*, Springer, 2019.
5. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2012.
6. S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset.," in *Twenty-Ninth IAAI Conference*, 2017.
7. D. Purves, G. Augustine, D. Fitzpatrick, W. Hall, A. LaMantia, J. McNamara, and S. Williams, *Neuroscience.* Sinauer Associates, 2004.
8. J. W. Britton, L. C. Frey, J. Hopp, P. Korb, M. Koubeissi, W. Lievens, E. Pestana-Knight, and E. L. St, *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants.* American Epilepsy Society, Chicago, 2016.
9. G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currentseeg, ecog, lfp and spikes," *Nature reviews neuroscience*, vol. 13, no. 6, p. 407, 2012.
10. M. X. Cohen, *Analyzing neural time series data: theory and practice.* MIT press, 2014.
11. R. W. Picard, *Affective computing.* MIT press, 2000.
12. M. Pantic and L. J. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1370–1390, 2003.
13. P. V. Rouast, M. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Transactions on Affective Computing*, 2019.

14. M. Abujelala, C. Abellanoza, A. Sharma, and F. Makedon, "Brain-ee: Brain enjoyment evaluation using commercial eeg headband," in *Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments*, p. 33, ACM, 2016.
15. P. A. Abhang and B. W. Gawali, "Correlation of eeg images and speech signals for emotion analysis," *British Journal of Applied Science & Technology*, vol. 10, no. 5, pp. 1–13, 2015.
16. A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, "High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice.," *Cerebral cortex (New York, NY: 1991)*, vol. 7, no. 4, pp. 374–385, 1997.
17. X. Zhang and D. Wu, "On the vulnerability of cnn classifiers in eeg-based bcis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 814–825, 2019.
18. X. Wang, M. Magno, L. Cavigelli, M. Mahmud, C. Cecchetto, S. Vassanelli, and L. Benini, "Embedded classification of local field potentials recorded from rat barrel cortex with implanted multi-electrode array," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, IEEE, 2018.
19. X. Wang, M. Magno, L. Cavigelli, M. Mahmud, C. Cecchetto, S. Vassanelli, and L. Benini, "Rat cortical layers classification extracting evoked local field potential images with implanted multi-electrode sensor," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, IEEE, 2018.
20. M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2063–2079, 2018.
21. P.-N. Tan, *Introduction to data mining.* Pearson Education India, 2018.
22. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016. http://www.deeplearningbook.org.
23. D. Zwillinger and S. Kokoska, *CRC Standard Probability and Statistics Tables and Formulae.* Chapman & Hall, 2000.
24. D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers.* John Wiley & Sons, 2010.
25. G. Strang, *Linear algebra and its applications.* Brooks Cole, 2006.
26. T. Y. Chiu, T. Leonard, and K.-W. Tsui, "The matrix-logarithmic covariance model," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 198–210, 1996.
27. C. Van Loan, *Computational frameworks for the fast Fourier transform*, vol. 10. Siam, 1992.
28. Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, pp. 396–404, 1990.
29. A. V. Oppenheim, A. S. Willsky, and S. Nawab, *Signals and Systems.* Prentice Hall, 1996.
30. D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642–3649, 2012.
31. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

32. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
33. F. Chollet *et al.*, "Keras." https://keras.io, 2015.
34. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv e-prints*, p. arXiv:1412.6980, Dec 2014.
35. J. J. Bird, D. R. Faria, L. J. Manso, A. Ekart, and C. D. Buckingham, "A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction," *Complexity*, vol. 2019, 2019.
36. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, pp. 1137–1143, 1995.
37. M. López-Ibáñez, J. Dubois-Lacoste, L. Pérez Cáceres, T. Stützle, and M. Birattari, "The irace package: Iterated racing for automatic algorithm configuration," *Operations Research Perspectives*, vol. 3, pp. 43–58, 2016.
38. E. K. Burke, M. Gendreau, M. Hyde, G. Kendall, G. Ochoa, E. Özcan, and R. Qu, "Hyper-heuristics: a survey of the state of the art," *Journal of the Operational Research Society*, vol. 64, no. 12, pp. 1695–1724, 2013.
39. A. Martín, R. Lara-Cabrera, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "Evodeep: A new evolutionary approach for automatic deep neural networks parametrisation," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 180–191, 2018.
40. F. Assunçao, N. Lourenço, P. Machado, and B. Ribeiro, "Denser: Deep evolutionary network structured representation," *arXiv preprint arXiv:1801.01563*, 2018.
41. J. J. Bird, A. Ekart, and D. R. Faria, "Evolutionary optimisation of fully connected artificial neural network topology," in *SAI Computing Conference 2019*, SAI, 2019.
42. D. C. Montgomery, *Design and Analysis of Experiments.* John Wiley & Sons, 8th ed. ed., 2012.
43. S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2013.